FR000121

US

Europäisches Patentamt | European Patent Office | Office européen des brevets

J1002 U.S. PTO
09/989252
11/20/01

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

00403281.9

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 22/08/01
LA HAYE, LE

EPA/EPO/OEB Form 1014 - 02.91

Europäisches Patentamt | European Patent Office | Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 00403281.9
Demande n°:

Anmeldetag:
Date of filing: 23/11/00
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Koninklijke Philips Electronics N.V.
5621 BA Eindhoven
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Coding method and associated decoding method

In Anspruch genommene Prioriät(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:
State:
Pays:

Tag:
Date:
Date:

Aktenzeichen:
File no.
Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

THIS PAGE BLANK (USPTO)

« CODING METHOD AND ASSOCIATED DECODING METHOD»

FIELD OF THE INVENTION

The present invention relates to the technical field of video encoders comprising base layer coding means, provided for receiving a video sequence and generating therefrom base layer signals that correspond to the video objects (VOs) contained in the video frames of said sequence and constitute a first bitstream suitable for transmission at a base layer bit rate to a video receiver, and enhancement layer coding means, provided for receiving said video sequence and a decoded version of said base layer signals and generating therefrom enhancement layer signals associated with corresponding ones of the compressed base layer video frames and suitable for transmission at an enhancement layer bit rate to said video receiver. More precisely, it relates to a method for coding the VOs of said sequence comprising the steps of :

(A) segmenting the video sequence into said VOs ;

(B) coding the successive video object planes (VOPs) of each of said VOs, said coding step itself comprising sub-steps of coding the texture and the shape of said VOPs, said texture coding sub-step itself comprising a motion compensated prediction operation subdivided into a zero value prediction for the VOPs called intracoded or I-VOPs, coded without any temporal reference to another VOP, an unidirectional prediction for the VOPs called predictive or P-VOPs, coded using only a past I- or P-VOP as a temporal reference, and a bidirectional prediction for the VOPs called bidirectional predictive or B-VOPs, coded using both past and future I- or P-VOPs as temporal references.

The invention also relates to computer executable process steps provided for carrying out such a coding method, and to a corresponding decoding method.

BACKGROUND OF THE INVENTION

The temporal scalability is a feature now offered by several video coding schemes, and it is, for example, one of the numerous options of the MPEG-4 video standard. A base layer is encoded at a given frame rate. Then an additional layer, called enhancement layer, is also encoded, in order to provide a higher temporal resolution at the display side. At the decoding side, only the base layer is usually decoded, but the decoder may also, in addition, decode the enhancement layer, which allows to output more frames per second.

Several structures are used in MPEG-4, and for example the video objects (VOs), which are the entities that a user is allowed to access and manipulate, and the video object planes (VOPs), which are instances of a video object at a given time. In an encoded bitstream, different types of VOPs can be found : intra coded VOPs, using only

spatial redundancy, predictive coded VOPs, using motion estimation and compensation from a past reference VOP, and bidirectionally predictive coded VOPs, using motion estimation and compensation from past and future reference VOPs. As the MPEG-4 video standard is a predictive coding scheme, some temporal references have to be defined for each coded non-intra VOP. In the single layer case or in the base layer of a scalable stream, temporal references are defined by the standard in a unique way. On the contrary, for the temporal enhancement layer of an MPEG-4 stream, three VOPs can be taken as a possible temporal reference for the motion prediction : the most recently decoded VOP of the enhancement layer, or the previous VOP in display order of the base layer, or the next VOP in display order of the base layer, as illustrated in Fig.1 where these three possible choices are shown for a P-VOP and a B-VOP of the temporal enhancement layer (each arrow corresponds to a possible temporal reference) : one reference has to be selected for each P-VOP of the enhancement layer and two for each B-VOP of the same layer.

Moreover, as a predictive coding scheme, scene-cut handling is a major feature for an MPEG-4 video encoder : when a scene-cut occurs, it is no longer possible to code the first VOP which immediately follows the scene-cut by predicting it from the preceding VOP, which is completely different from it. In case of temporally scalable encoding, the problem is even more complex, since the scene-cut may occur between two VOPs of the enhancement layer while having still to be handled on the base layer.

It must also be noted that, under certain conditions, there is a large difference of quality between the displayed images of the base layer and those of the enhancement layer, for example when the available bandwidth for each layer is very different. In that case, the subjective quality of the decoded sequence can be quite low because of the flickering effect, even if only a few frames (those of the base layer) have a significantly lower quality compared with the average of the sequence.

SUMMARY OF THE INVENTION

It is therefore a first object of the invention to propose a predictive coding scheme using an improved temporal distance criterion for the selection of the temporal references.

To this end the invention relates to a coding method such as defined in the introductory paragraph of the description and in which the temporal references of the enhancement layer P-VOPs or B-VOPs are selected only as the temporally closest candidates or the two ones respectively, without any consideration of the layer they belong to.

A further object of the invention is to propose a predictive coding scheme in which a particular processing allows, for said selection of the temporal references, to

solve the problem of scene-cuts that may occur between two VOPs of the enhancement layer.

To this end, the invention relates to a coding method such as defined in the introductory paragraph of the description and in which the temporal references of the enhancement layer VOPs are selected, when a scene cut occurs and said enhancement layer VOPs are located between the last base layer VOP of a scene and the first base layer VOP of the following scene, according to the following specific processing :

(a) VOPs located before the scene cut : no constraint is applied to the coding type, and the use of the next VOP in display order of the base layer as a temporal reference is forbidden ;

(b) the VOP located just immediately after the scene cut : P coding time is enforced with the next VOP in display order of the base layer as a temporal reference ;

(c) other VOPs located after the scene cut : no constraint is applied to the coding type, and the use of the previous VOP in display order of the base layer as a temporal reference is forbidden.

A further object of the invention is also to propose a corresponding decoding scheme in which the problem of the large difference of quality between the displayed images of the base and enhancement layers is solved.

To this end, the invention relates to a decoding method for processing signals that have been coded according to one of said embodiments of the coding method according to the invention and in which the poor quality images of the base layer are replaced by images interpolated on the basis of the preceding and following images of the enhancement layer.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described, by way of example, with reference to the accompanying drawings in which :

Fig.1 illustrates the possible temporal references in the case of a scalable MPEG-4 video stream ;

Fig.2 illustrates, according to an embodiment of the invention, the process applied to VOPs in case of a scene-cut occurring between two VOPs.

DETAILED DESCRIPTION OF THE INVENTION

As seen above, during the encoding process, one reference has to be selected (out of three candidates) for each P-VOP of the enhancement layer and two for each B-VOP of said layer. It is then decided to select these temporal references of the enhancement VOPs only on the basis of a temporal distance criterion, without any consideration of the layer they belong to. Consequently, for a P-VOP the reference is the

temporally closest candidate, and for a B-VOP the references are the two temporally closest candidates.

However, when a scene-cut occurs, there is no possible choice of the temporal reference in the base layer, and it is decided to code as an I-VOP the first VOP of the base layer after the scene-cut. On the enhancement layer, such an intra coding is not used, it is only needed to ensure that there is no inter-scene prediction, which is obtained by carefully choosing the coding type (P or B) and the corresponding temporal references of the VOPs.

The following specific processing is then applied to all enhancement layer VOPs located between the last base layer VOP of a scene and the first base layer VOP of the following scene :

(a) VOPs located before the scene cut : no constraint is applied to the coding type, and the use of the next VOP in display order of the base layer as a temporal reference is forbidden ;

(b) the VOP located just immediately after the scene cut :
P coding time is enforced with the next VOP in display order of the base layer as a temporal reference ;

(c) other VOPs located after the scene cut : no constraint is applied to the coding type, and the use of the previous VOP in display order of the base layer as a temporal reference is forbidden.

These three situations are illustrated in Fig.2. By comparing Figs.1 and 2, it is clearly seen that the conditions (a), i.e. no use of a next VOP for a VOP located before the scene-cut, (b), i.e. the next VOP of the base layer as a temporal reference, and (c), i.e no previous VOP of the base layer as a temporal reference, are satisfied.

The VOPs thus coded are transmitted and/or stored, and later received by a decoder, in order to be decoded and displayed. Under certain conditions, there may have a large difference of quality between the images of the base layer and those of the enhancement layer, for example when the available bandwidth for each layer is very different. In that case, the subjective quality of the displayed, decoded sequence will be low, owing to a flickering effect, even if only a few frames in the base layer have a significantly lower quality compared with the average quality of the sequence. This drawback may be avoided if said poor quality images of the base layer are not displayed and are replaced by images interpolated on the basis of the preceding and following images of the enhancement layer.

CLAIMS :

1. For use in a video encoder comprising base layer coding means, provided for receiving a video sequence and generating therefrom base layer signals that correspond to the video objects (V0s) contained in the video frames of said sequence and constitute a first bitstream suitable for transmission at a base layer bit rate to a video receiver, and enhancement layer coding means, provided for receiving said video sequence and a decoded version of said base layer signals and generating therefrom enhancement layer signals associated with corresponding ones of the compressed base layer video frames and suitable for transmission at an enhancement layer bit rate to said video receiver, a method for coding the V0s of said sequence comprising the steps of :

(A) segmenting the video sequence into said V0s ;

(B) coding the successive video object planes (VOPs) of each of said VOs, said coding step itself comprising sub-steps of coding the texture and the shape of said VOPs, said texture coding sub-step itself comprising a motion compensated prediction operation subdivided into a zero value prediction for the VOPs called intracoded or I-VOPs, coded without any temporal reference to another VOP, an unidirectional prediction for the VOPs called predictive or P-VOPs, coded using only a past I- or P-VOP as a temporal reference, and a bidirectional prediction for the VOPs called bidirectional predictive or B-VOPs, coded using both past and future I- or P-VOPs as temporal references, the temporal references of the enhancement layer P-VOPs or B-VOPs being selected only as the temporally closest candidates or the two ones respectively, without any consideration of the layer they belong to.

2. For use in a video encoder comprising base layer coding means, provided for receiving a video sequence and generating therefrom base layer signals that correspond to the video objects (V0s) contained in the video frames of said sequence and constitute a first bitstream suitable for transmission at a base layer bit rate to a video receiver, and enhancement layer coding means, provided for receiving said video sequence and a decoded version of said base layer signals and generating therefrom enhancement layer signals associated with corresponding ones of the compressed base layer video frames and suitable for transmission at an enhancement layer bit rate to said video receiver, a method for coding the V0s of said sequence comprising the steps of :

(A) segmenting the video sequence into said V0s ;

(B) coding the successive video object planes (VOPs) of each of said V0s, said coding step itself comprising sub-steps of coding the texture and the shape of said VOPs, said texture coding sub-step itself comprising a motion compensated prediction operation subdivided into a zero value prediction for the VOPs called intracoded or I-VOPs, coded without any temporal reference to another VOP, an unidirectional prediction for the VOPs called predictive or P-VOPs, coded using only a past I- or P-VOP as a temporal reference, and a bidirectional prediction for the VOPs called bidirectional

predictive or B-VOPs, coded using both past and future I- or P-VOPs as temporal references, the temporal references of the enhancement layer VOPs being selected, when a scene cut occurs and said enhancement layer VOPs are located between the last base layer VOP of a scene and the first base layer VOP of the following scene, according to the following specific processing :

(a) VOPs located before the scene cut : no constraint is applied to the coding type, and the use of the next VOP in display order of the base layer as a temporal reference is forbidden ;

(b) the VOP located just immediately after the scene cut : P coding time is enforced with the next VOP in display order of the base layer as a temporal reference ;

(c) other VOPs located after the scene cut : no constraint is applied to the coding type, and the use of the previous VOP in display order of the base layer as a temporal reference is forbidden.

3. Computer executable process steps stored on a computer readable medium and provided for carrying out a coding method according to anyone of claims 1 and 2.

4. A decoding method for processing signals that have been coded according to the coding method as claimed in anyone of claims 1 and 2, wherein the poor quality images of the base layer are replaced by images interpolated on the basis of the preceding and following images of the enhancement layer.

Abstract

The invention relates, for use in a video encoder with base layer coding means and enhancement layer coding means, to a method of coding the video objects (VOs) of a sequence according to the following steps : segmentation of the sequence, and coding operation of the texture and shape of said VOs. According to a preferred embodiment, the texture coding operation itself comprises motion compensated prediction operations, during which the temporal references of the enhancement layer VO planes (VOPs) of type P or B are selected only as the two temporally closest candidates or the two ones respectively, without any consideration of the layer they belong to.

THIS PAGE BLANK (USPTO)

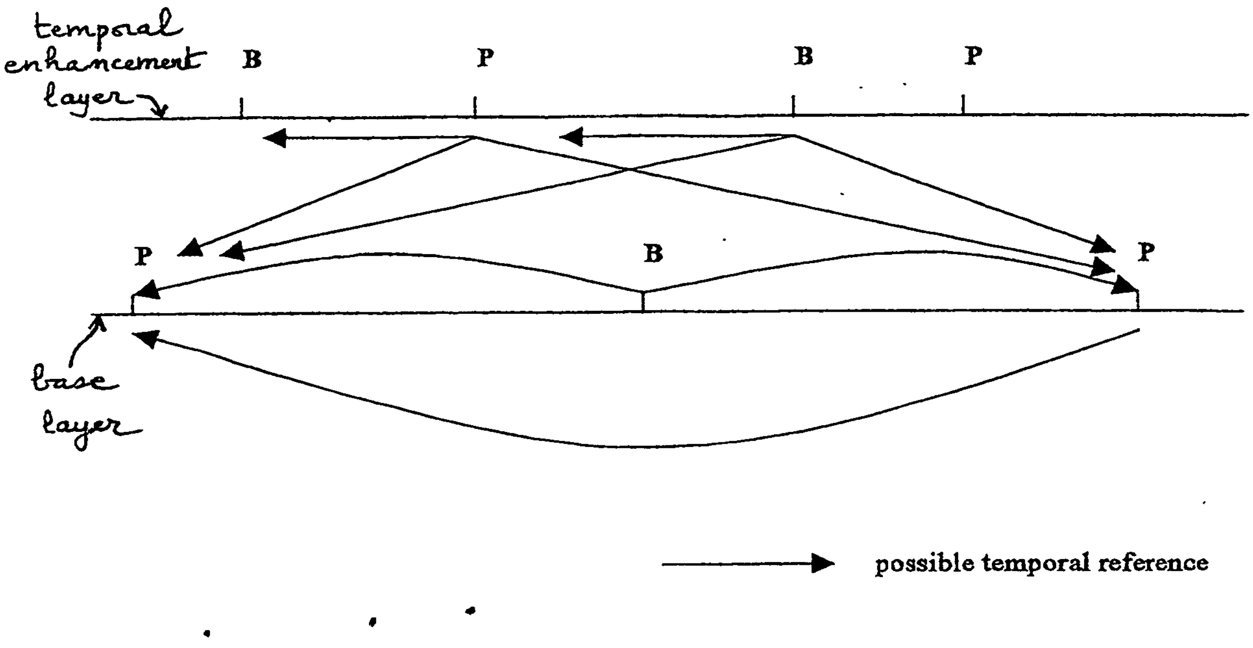

figure 1 : temporal references in a scalable MPEG-4 video stream

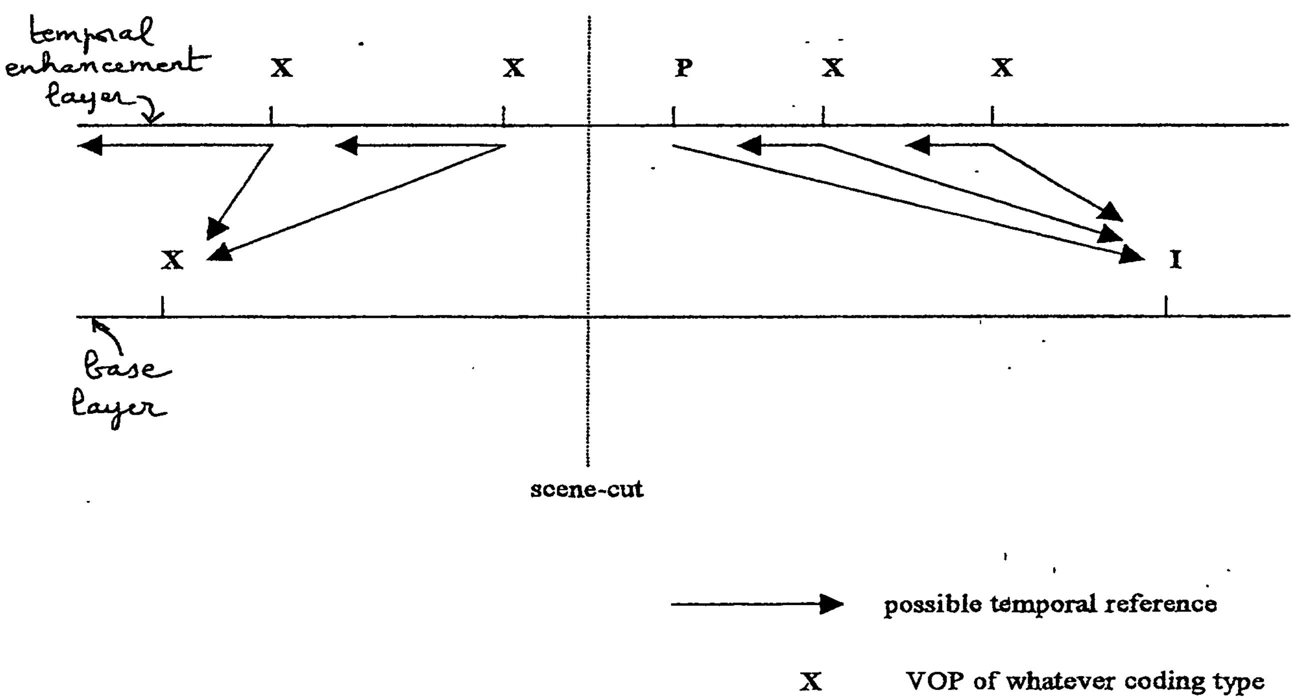

figure 2: VOP coding type and temporal reference around a scene-cut